



VERSANCE.AI

WHITEPAPER

# A (LLM) LARGE LANGUAGE MODEL FINANCIAL CONTENT (FINCON) FRAMEWORK

2024

---

# Abstract

Versance.ai is a pioneering AI-based FinCon (financial content) business that leverages the power of large language models (LLMs), vector databases and prompt engineering to provide innovative AI solutions to public companies and investors. This white paper introduces the company's unique approach to harnessing AI, focusing on creating a financial data and content framework to provide a suite of services to public companies across all sectors and markets, while leveraging Versance.ai's existing content datasets, aggregation systems, websites and expertise.

The company's suite of applications and services includes **Investor Relations GPT chatbot**, a **press release authoring assistant**, **Versance.ai Legalbot** securities regulations and legal requirements vetting and review, an **article writing assistant**, automatic **article publishing** to one of Versance.ai's investor websites, and a **social media authoring** tool. These services are designed to streamline and enhance investor communications, education, and information dissemination for public companies and investors.

## 1. Introduction

The rapid growth and evolution of artificial intelligence have led to the development of large language models (LLMs), resulting in a significant shift in the field of natural language processing across various domains. This change has sparked interest in the potential application of these models in the financial sector. However, it is clear that the acquisition and management of high-quality, relevant, and up-to-date data is a critical factor in the development of an effective and efficient financial language model.

Applying language models in the financial arena presents complex challenges. These range from difficulties in obtaining data, dealing with diverse data formats and types, managing data quality inconsistencies, to the essential requirement of up-to-date information and regulatory compliance. Particularly, the extraction of historical or specialized financial information can be complex due to varying data mediums such as web platforms, APIs, PDF documents, and images.

In response to these challenges, Versance.ai introduces an end-to-end framework for AI-based FinCon (financial content) that leverages the power of large language models (LLMs), vector databases and prompt engineering. Adopting a data-centric approach and utilizing the company's datasets and content pipelines, Versance.ai emphasizes the crucial role of data acquisition, cleaning, and preprocessing in developing AI-based FinCon models.

The Versance.ai FinCon Framework is summarized as follows:

- **Data-centric approach.** Recognizing the significance of data curation, Versance.ai adopts a data-centric approach and implements rigorous cleaning and preprocessing methods for handling its proprietary datasets, data aggregation systems, and other data sources of varied data formats and types, thereby ensuring high-quality data.

- **End-to-end framework:** Versance.ai embraces a full-stack framework with four layers:
  - **Data source layer: This layer assures comprehensive markets and investing coverage, addressing the temporal sensitivity of financial data and information through real-time capture.**
  - **Data engineering layer: Primed for real-time NLP data processing, this layer tackles the inherent challenges of high temporal sensitivity and low signal-to-noise ratio in financial data.**
  - **LLMs layer: Focusing on a range of fine-tuning methodologies, this layer mitigates the highly dynamic nature of financial data, ensuring the model's relevance and accuracy.**
  - **Prompt engineering layer: Leveraging tools like LangChain and Pinecone, this layer optimizes interactions with the LLM. LangChain manages data within model chains, while Pinecone captures semantic text meanings. Together, they assure precise, contextually accurate AI responses, enhancing overall system performance.**
  - **Application layer: A group of practical applications and demos, this layer highlights the capability of Versance.ai to transform high cost, high value aspects of investor marketing and communications for public companies, while creating immense value for investors.**

Versance.ai's vision is to empower public companies and transform and improve communications and investor marketing by deploying powerful AI based solutions.

AI offers the promise of true one to one marketing that delivers immense value to all participants. Versance.ai's data driven (LLM) large language model financial content (FinCon) framework is the foundation for a suite of services never before possible.

### LLMs and ChatGPT

Large Language Models (LLMs) like ChatGPT are transforming the way businesses operate, providing a powerful tool for generating human-like text. Versance.ai leverages these models, combined with its proprietary data management, regulatory compliance framework, and prompt engineering techniques, to deliver a range of AI-powered services.

### LLMs in Financial Data

LLMs have significant potential in the financial information sector, where they can be used to analyze complex financial data, generate insightful reports, and provide real-time assistance to investors and companies. Versance.ai is harnessing this potential to deliver a suite of services that enhance the efficiency and effectiveness of financial communication and decision-making.

## 2. Data-Centric Approach

Versance.ai's approach is centered on the effective management and utilization of data. The company's AI platform manages multiple data sources, using vector databases and prompt engineering to enhance the capabilities of LLMs. This data-centric approach allows Versance.ai to deliver highly accurate and relevant AI-powered services.

### Financial & Public Company Data and Unique Characteristics

Financial and public company data has several unique characteristics that make it particularly suited to analysis by AI:

- It is highly structured, with clear relationships between different data points.
- It is time-sensitive, with the relevance of data often changing rapidly.
- It is highly regulated, with strict rules governing how financial and marketing information can be used and shared.
- It is complex, with a wide range of factors influencing outcomes.

Versance.ai's AI platform is designed to handle these unique characteristics, delivering accurate, regulatory compliant and timely solutions.

### Challenges in Handling Financial Data

Despite its potential, financial data presents several challenges. These include the need for accurate and timely data, the complexity of financial markets and regulations, and the strict regulatory environment governing public companies and financial information. Versance.ai's AI platform is designed to overcome these challenges, using advanced AI techniques to analyze and interpret financial data effectively.

### Financial Data and Unique Characteristics

**Real Time Financial data** comes from a variety of sources, with unique characteristics. We delve into the specifics of different financial data sources, such as Financial News, Company Filings and Company Announcements.

**Financial news** carries vital information about the world economy, specific industries, and individual companies. This data source typically features:

- *Timeliness: Financial news reports are timely and up-to-date, often capturing the most recent developments in the financial world.*
- *Dynamism: The information contained in financial news is dynamic, changing rapidly in response to evolving economic conditions and market sentiment.*
- *Influence: Financial news has a significant impact on financial markets and share prices, influencing traders' decisions and potentially leading to dramatic market movements.*

**Company filings and announcements** are official documents that corporations submit to regulatory bodies, providing insight into a company's financial health and strategic direction. They feature:

- **Granularity: These documents offer granular information about a company's; financial status, including assets, liabilities, revenue, and profitability; project and development status; and management and business status**
- **Reliability: Company filings contain reliable and verified data vetted by regulatory bodies.**
- **Periodicity: Company filings are periodic, usually submitted on a quarterly or annual basis, offering regular snapshots of a company's financial situation.**
- **Impactfulness: Company announcements often have substantial impacts on the market, influencing stock prices and investor sentiment.**

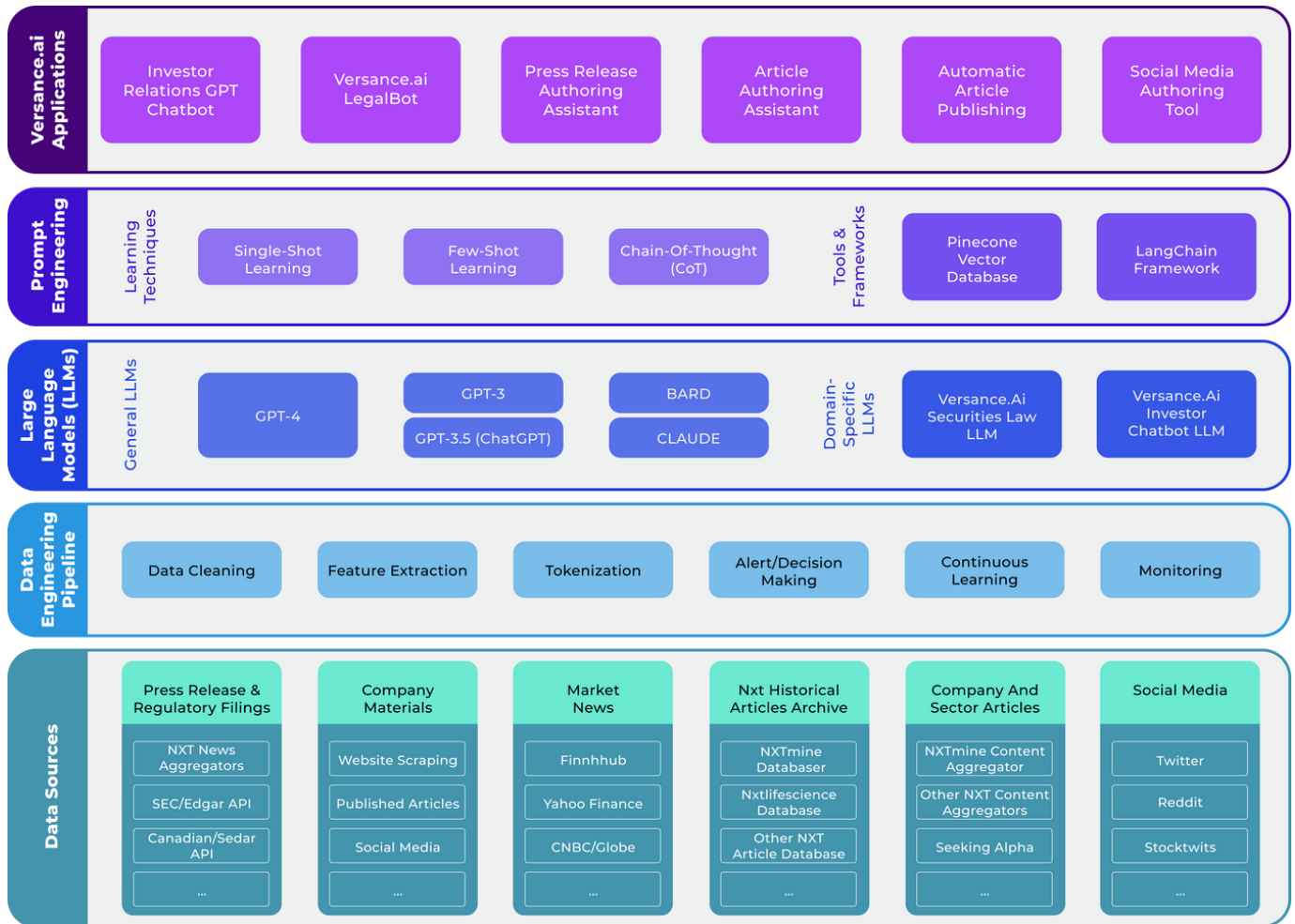


Figure 1: Versance.ai Framework

### 3. Overview of Versance.ai: A Financial Information Platform

Versance.ai's AI platform is a robust and flexible platform for the development of AI-powered financial services. Key features of Versance.ai include:

- A data-centric approach, with a focus on managing and utilizing financial and public company data effectively.
- The use of vector databases and prompt engineering to enhance the capabilities of LLMs.
- A focus on regulatory compliance and review
- The ability to handle multiple data sources, including press releases, news, historical

#### Data Sources

Versance.ai's AI platform manages multiple data sources, each playing a crucial role in the services provided:

- **Press Releases & Regulatory Filings:** These are a primary source of information about a company's activities, financial performance, and strategic direction. They provide the most timely and relevant data.
- **Company Materials:** Public companies produce a wide range of material that explain their business and strategy. Websites, investor decks, social media, videos, and interviews are a primary, vetted source of information

- **Market News:** Live market news and articles, and market data feeds from reputable sources providing real time context on economic, business and company activities, trends, and the markets in which they operate.
- **NXT Historical Articles Archive:** A content and data resource with a database of over 100,000 articles aggregated over the last four years. Focused on public companies, markets, economics and multiple sectors such as mining and life sciences.
- **Aggregated Company and Sector Articles:** NXT aggregates articles from over 400 independent sources daily. These articles provide information about companies and the sectors in which they operate. They offer insights into market trends, competitive dynamics, regulatory changes, and other factors that may impact a company's performance.
- **Social Media:** Investors use social media like Twitter, Reddit and Stocktwits as a primary source of information and opinion.
- **Data Cleaning:** Real-time data can be noisy and inconsistent. Therefore, real-time data cleaning involves removing irrelevant data, handling missing values, text normalization (like lowercasing), and error corrections.
- **Tokenization:** In real-time applications, tokenization has to be performed on the fly. This involves breaking down the stream of text into smaller units or tokens.
- **Stop Word Removal and Stemming / Lemmatization:** For real-time processing, a predefined list of stop words can be used to filter out these common words from the stream of tokens. Likewise, stemming and lemmatization techniques can be applied to reduce words to their root form.
- **Feature Extraction and Sentiment Analysis:** Feature extraction involves transforming raw data into an input that can be understood by machine learning models. Sentiment analysis can also be performed on the cleaned data. This is where we categorize a span of text as positive, negative, or neutral.
- **Prompt Engineering:** The creation of effective prompts, and prompt chains that can guide the language model's generation process as it produces customized content in desired formats and forms.
- **Alerts/Decision Making:** Once the prompt is entered, the results need to be communicated or acted upon. This might involve triggering alerts based on certain conditions, informing real-time decision-making processes, or feeding the output into another system.
- **Continuous Learning.** In real-time systems, the models should adapt to changes in the data. Continuous learning systems can be implemented, where models are periodically retrained on new data or online learning algorithms are used that can update the model with each new data point.

## Data Engineering Pipeline

Financial markets operate in real-time and public companies issue news and are newsworthy on a daily basis. Prices of securities and investor sentiment can change rapidly in response to new information, and delays in processing that information can result in missed opportunities. The primary challenge with a NLP pipeline is managing and processing the continuous inflow of data efficiently. The first step in the pipeline is to set up a system to ingest data. This data could be streaming from various data sources.

Below are the steps to design a NLP pipeline for data ingestion.

- *Monitoring: Real-time systems require continuous monitoring to ensure they are functioning correctly. Any delays or issues in the pipeline can have immediate impacts, so it's important to have robust monitoring and alerting in place.*

## Large Language Models (LLMs)

LLMs are a critical component of Versance.ai's AI platform. These models are designed to understand and generate human-like text, making them highly effective tools for a variety of applications.

### General LLMs

General LLMs like GPT4 are trained on a diverse range of internet text. However, they can be prompted on specific tasks or domains to improve their performance. These models are capable of understanding context, generating coherent responses, and even exhibiting a degree of creativity in their outputs. They are used in a variety of applications, from writing reports, drafting emails to writing code.

Once the data stream has been properly prepared, it is used with LLMs to generate content such as news releases and articles, chat responses, and financial analysis. The LLM layer includes:

- LLM APIs: APIs from established LLMs provide baseline language capability.
- Trainable models: Trainable models allow users to fine-tune on private data, customizing for financial applications.
- Fine-tuning methods: Various fine-tuning methods allow Versance.ai to adapt LLMs to custom applications.
- Prompt engineering: Various tools, techniques

and applications are used to create a series of instructions that guide the LLMs

### Domain-Specific LLMs

In addition to using general LLMs, Versance.ai will also build domain-specific LLMs. These models will be trained on a specific set of data related to a particular topic or industry, such as securities law, biotech or mining content. By focusing on a specific domain, these models can provide more accurate and relevant outputs for tasks within that domain.

The combination of general and domain-specific LLMs allows Versance.ai to provide a wide range of services tailored to the needs of public companies and investors. For example, a domain-specific LLM trained on securities law can be used to create a securities regulations and legal requirements vetting tool, while a general LLM like GPT4 can be used to generate human-like text for an investor relations chatbot.

### Fine-tuning an Open Source Pre-trained LLM Model

Fine-tuning is a process that involves adapting a pre-trained model to a specific task or domain by training it further on a smaller, task-specific dataset

### Advantages of Using Open Source Pre-trained LLM's

- **Cost-Effective:** By utilizing open-source pre-trained models, usage fees related to the compute resources needed for training are reduced.
- **Data Privacy:** Control over data is assured since data is not mixed or transmitted through external API's.
- **Flexibility:** Control and flexibility to customize the model and the training process

## Prompt Engineering

Prompt engineering is the process of refining interactions with AI systems, such as ChatGPT, to produce optimal responses. By crafting effective queries or inputs – referred to as prompts – an AI language model can be guided towards generating desired responses.

Prompt engineering becomes even more powerful when combined with additional technologies like vector databases, such as Pinecone, and prompt frameworks, such as LangChain, which allow higher precision in the AI's responses.

LangChain enhances AI interactions by managing data within language model chains. For example, LangChain can efficiently condense long-form content or complex tasks to fit within the LLM context window (word limit per prompt). Used in tandem with vector databases, such as Pinecone, which augment the LLM, we can generate more precise and contextually accurate AI responses through sophisticated prompt engineering.

## 4. Versance.ai Applications

Versance.ai's AI platform will be used to deliver a suite of services, including:

- **Investor Relations GPT:** a "smart" interactive custom chatbot for interacting with shareholders and investors. Programmed to be regulatory compliant and able to answer detailed questions accurately.

The Investor Relations GPT chatbot can be further customized to identify shareholders and provide customized responses and information.

Our approach involves using a large dataset of company materials to train a machine learning model. Data will be captured from a variety of

sources such as company news and filings, all investor marketing materials, website content, management interviews, webinars, articles, sector research and analyst commentary,

For indexing and cross-referencing, techniques like Named Entity Recognition (NER) can identify and categorize key information in our dataset, such as company names, financial terms, and legal requirements to build a structured knowledge base for the chatbot.

- **Versance.ai LegalBot** provides securities regulations review and legal vetting of content generation within Versance.ai's suite of solutions. It's also able to proofread any written content and generate comments and provide suggestions on compliance with Canadian, US and international securities laws.

Starting with a rich, well-structured dataset containing securities regulations, including analysis of various related aspects, such as the scope of rules and corporate governance requirements in Canadian and US securities regulation.

The service will be used in three ways:

- During generation: integrated into Versance.ai's applications for marshaling content generated by LLMs and humans. This is achieved through machine learning NLP vector databases for semantic searching and/or rule-based systems.
- During vetting: when used to review and vet written content for regulatory and legal compliance.
- During LLM Fine-tuning: As part of the datasets that make up the corpus used for the fine-tuning of pre-trained LLMs used in applications. Fine-tuning a pre-trained model like MossiacML's MPT-7B on our specific



dataset will help the chatbot better understand the context and nuances of investor relations and regulatory compliance.

- **Press Release Authoring Assistant** for public companies, initially focused on the mining sector then expanding to all sectors.

Our Press Release Authoring Assistant utilizes advanced AI and NLP to generate press releases based on key input data, such as the nature of the news, the involved parties, financial details, and more. It's trained to understand the nuances of corporate communications, ensuring that the tone, style, and content of each release are perfectly tailored to the audience and the news at hand. Multi-language support allows for the generation of press releases in various languages, ensuring a global reach. The system utilizes the Versance.ai LegalBot to ensure press release content is compliant with all regulatory and legal standards.

- **Article Authoring Assistant**, which uses a range of data sources including press releases, to generate insightful articles that provide value to investors.

Our Article Authoring & Publishing tool combines cutting-edge NLP and machine learning technologies to generate high-quality articles based on a variety of input parameters. The tool can draft a wide range of content including investor articles for publication, corporate blog posts, data-driven market analysis pieces, and more, all while ensuring strict adherence to regulatory guidelines.

- **Automatic Article Publishing** provides clients with the ability to publish articles directly to one of Versance.ai's investor websites such as NXTmine.com or NXTlifescience.com.

- **Social Media Authoring** tool produces a range of posts with photos or graphics, using articles, press releases, market news, and relevant sector and business articles as sources.

Our Social Media Authoring & Production tool can generate a range of content, from short tweets to detailed LinkedIn posts, based on user-defined input parameters. All content generated is compliant with relevant social media guidelines and regulatory standards.

### Potential Partnerships

Versance.ai is building a platform capable of providing a wide range of financial services for public companies and investors. We intend to seek partners who are interested in utilizing the platform to create complimentary services such as Robo Advisors, Securities Law Advisors, Market Sentiment Analysis, and Trading Analytics Systems.

## 5. Conclusion

Versance.ai is at the forefront of the application of AI in the financial content sector, leveraging a data-centric approach and an end-to-end framework to deliver innovative solutions. Versance.ai is innovating by leveraging pre-existing LLMs and fine-tuning them to specific FinCon financial applications. This approach significantly reduces adaptation costs and computational requirements while providing a deeper understanding, greater security and flexibility. The company's suite of services which includes an **Investor Relations GPT chatbot**, a **press release authoring** assistant, the Versance.ai Legalbot securities regulations and legal requirements vetting and review system, an **article writing assistant**, automatic **article publishing** and marketing to one of Versance.ai's investor websites, and a **social media authoring** tool, will transform the way public companies and investors communicate and make decisions.

## Important Notice

This white paper is for informational purposes only and does not constitute an offer to sell, a solicitation to buy, or an endorsement, recommendation, or sponsorship of any company or security by Versance Technologies Inc. or any other party. The content of this white paper is provided for general information only and is subject to change without notice.

The information contained within this document is based on sources considered to be reliable, but its accuracy and completeness are not guaranteed. It should not be relied upon as a basis for making investment decisions. The views, thoughts, and opinions expressed in this white paper belong solely to [Versance.ai](https://Versance.ai) and do not necessarily reflect the official policy or position of any other agency, organization, employer, or company.

[Versance.ai](https://Versance.ai) makes no representations or warranties, express or implied, as to the accuracy, completeness, or reliability of the information contained in this white paper. The technologies, techniques, and applications described within may be subject to intellectual property rights reserved by Versance Technologies Inc. or other third parties. No use of such information is authorized hereunder except as expressly set forth in written agreements with Versance Technologies Inc.

Readers are cautioned to conduct their own due diligence and consult with their professional advisors about the material contained in this white paper. Versance Technologies Inc. assumes no liability for any loss or damage of any kind arising out of the use of all or any part of this material.

This document may contain forward-looking statements, including statements regarding the development of services, technological advancements, and future plans of Versance Technologies Inc. Such forward-looking statements are based on current expectations and involve inherent risks and uncertainties, including factors that could delay, divert, or change any of them, and could cause actual outcomes and results to differ materially from current expectations.

No regulatory authority has reviewed or approved of any of the material set forth in this white paper. The publication, distribution, or dissemination of this white paper does not imply that applicable laws, regulatory requirements, or rules have been complied with.

